Subsurface Imaging with Support Vector Machines

Brendt Wohlberg, Daniel M. Tartakovsky, and Alberto Guadagnini

LAUR-05-0493

***Abstract*— A typical subsurface environment is heterogeneous, consists of multiple materials (geologic facies), and is often insufficiently characterized by data. The ability to delineate geologic facies and to estimate their properties from sparse data is essential for modeling physical and biochemical processes occurring in the subsurface. We demonstrate that the Kernel Support Vector Machine is a viable and efficient tool for facies delineation, and contrast it with existing geostatistical approaches. To illustrate our approach, and to demonstrate its advantages, we construct a synthetic porous medium consisting of two heterogeneous materials and then estimate boundaries between these materials from a few selected data points. We also introduce and analyze the use of regression Support Vector Machines to estimate the parameter values between point where the parameter is sampled. Our analysis shows that the error in facies delineation by means of Support Vector Machines decreases logarithmically with increasing sampling density.**

***Index Terms*— Support Vector Machine, Machine Learning, geostatistics, geologic facies, data analysis**

I. Introduction

Our knowledge of the spatial distribution of the physical properties of geologic formations is often uncertain because of ubiquitous heterogeneity and the sparsity of data. Geostatistics has become an invaluable tool for estimating such properties at points in a computational domain where data are not available, as well as for quantifying the corresponding uncertainty. Geostatistical frameworks treat a formation's properties, such as hydraulic conductivity $K(\mathbf{x})$, as random fields that are characterized by multivariate probability density functions or, equivalently, by their joint ensemble moments. Thus, $K(\mathbf{x})$ is assumed to vary not only across the physical space (coordinate $\mathbf{x}$), but also in probability space (this variation may be represented by another coordinate $\boldsymbol{\xi}$, which is usually suppressed to simplify notation). Whereas spatial moments of K are obtained by sampling $K(\mathbf{x})$ in physical space (across $\mathbf{x}$), its ensemble moments are defined in terms of samples collected in probability space (across $\boldsymbol{\xi}$). Since in reality only a single realization of a geologic site exists, it is necessary to invoke the ergodicity hypothesis in order to substitute the sample spatial statistics, which can be calculated, for the ensemble statistics, which are actually required. Ergodicity cannot be proved and requires a number of modeling assumptions, e.g., [1, Sec. 2.7 and references therein]. One of the most popular geostatistical approaches to facies delineation employs discontinuous geostatistical models, such as the Indicator Kriging (IK) [2], [3], [4]. IK has also found its way into image processing [5].

D. M. Tartakovsky and B. Wohlberg, Theoretical Division, Group T-7, MS B284, Los Alamos National Laboratory, Los Alamos, NM 87545. (e-mail: *dmt@lanl.gov*; *brendt@t7.lanl.gov*)

A. Guadagnini, D.I.I.A.R., Politecnico di Milano, Piazza L. Da Vinci, 3220133 Milano, Italy. (e-mail: *alberto.guadagnini@polimi.it*)

Machine learning provides an alternative to the geostatistical framework by allowing one to make predictions in the absence of sufficient data parameterization, without treating geologic parameters as random and, hence, without the need for the ergodicity assumptions. Intimately connected to the field of pattern recognition, machine learning refers to a family of computational algorithms for data analysis that are designed to automatically tune themselves in response to data. Neural networks [6] are an example of such a class of algorithms that has found its way into hydrologic modeling. While versatile and efficient for many important applications, such as the delineation of geologic facies [7], the theory of neural networks remains to a large extent empirical in this context.

Recently, we [8] introduced another subset of the machine learning techniques — the Support Vector Machine (SVM) and its mathematical underpinning, the Statistical Learning Theory (SLT) of Vapnik [9]. While similar to neural networks in its goals, the SVM is firmly grounded in rigorous mathematical analysis, which allows one not only to assess its performance but to bound the corresponding errors as well. Like other machine learning techniques, the SVM and SLT enable one to treat the subsurface environment and its parameters as deterministic. Uncertainty associated with insufficient data parameterization is then represented and quantified by treating sampling locations as a random subset of all possible measurement locations. Such a formulation is ideally suited for subsurface imaging.

In [8], we used SVMs to locate a boundary between two materials in a perfectly stratified geologic formation. Such a boundary is by definition either a straight line (in two dimensions) or a plane (in three dimensions), so that available data are always linearly separable. Here we use a generalization of SVMs, which are known as Kernel SVMs, to delineate highly irregular boundaries between two heterogeneous geologic facies from a sparsely sampled parameter.

We formulate the problem of facies delineation in Section II. Section III provides a brief description of the general theory of Kernel Support Vector Machines, with an emphasis on their application in subsurface imaging. Kernel SVMs are then used in Section IV to reconstruct a boundary between two heterogeneous geologic facies from a few data points extracted from a randomly generated porous medium. In Section V we introduce a regression SVM, and propose its sequential generalization, to estimate parameter values at points where parameter data are not available. The last two sections also contrast the performance of the Kernel SVMs and the regression SVMs with that of a geostatistical approach.

II. A Problem of Facies Delineation

Consider the problem of reconstructing a boundary between two heterogeneous materials (geologic facies) from parameter data, say hydraulic conductivity or electric resistivity measurements $K_i = K(\mathbf{x}_i)$, collected at N selected locations $\mathbf{x}_i = (x_i, y_i)^T$, where $i \in \{1, \ldots, N\}$. The first step to facies delineation consists of analyzing a data histogram to assign to each data point a value of the indicator function,

$$I(\mathbf{x}_i) = \begin{cases} 1 & \mathbf{x}_i \in M_1 \\ 0 & \mathbf{x}_i \in M_2, \end{cases} \tag{1}$$

where M_1 and M_2 are the two facies.

Let $\mathcal{I}(\mathbf{x}, \boldsymbol{\alpha})$ be an estimate of a "true" indicator field $I(\mathbf{x})$, whose adjustable parameters $\boldsymbol{\alpha}$ are consistent with, and determined from, the available data $\{\mathbf{x}_i, I(\mathbf{x}_i)\}_{i=1}^N$. One would like to construct an estimate that is as close to the true field as possible, i.e., to minimize the difference between the two, $||I - \mathcal{I}||$. In general, both the indicator field I and the choice of sampling locations $\{\mathbf{x}_i\}_{i=1}^N$ can be modeled as random, and be described by a joint probability distribution $P(I, \mathbf{x})$ or, equivalently, a joint probability density function $p(I, \mathbf{x})$. Then the problem of obtaining the best estimate of the indicator field is equivalent to minimizing the functional

$$R = \int ||I - \mathcal{I}|| dP(I, \mathbf{x}) = \int ||I - \mathcal{I}|| p(I, \mathbf{x}) dI d\mathbf{x}. \tag{2}$$

Unfortunately, since in reality only a single geologic formation exists, there is no direct way to evaluate $P(I, \mathbf{x})$. Geostatistical and Statistical Learning techniques provide two alternatives for evaluating (2).

Geostatistical approaches use the L^2 norm in (2), and treat

1) the indicator function I is a random field, and
2) the choice of sampling locations $\{\mathbf{x}_i\}_{i=1}^N$ as deterministic.

Then the problem of minimizing (2) reduces to the minimization of the indicator variance

$$\sigma_I^2 = \int (I - \mathcal{I})^2 dP(I) = \int (I - \mathcal{I})^2 p(I) dI. \tag{3}$$

To approximate $p(I)$, geostatistical approaches assume ergodicity, i.e. that the sample statistics of I, such as mean μ_I, variance σ_I^2, and correlation function ρ_I computed from spatially distributed data $\{I(\mathbf{x}_i)\}_{i=1}^N$ can be substituted for the ensemble statistics. Furthermore, it is necessary to assume that these sampling statistics are representative of the whole field.

SLT [9] often uses the L^1 norm in (2), and treats

1) the indicator function I as deterministic, and
2) the choice of sampling locations $\{\mathbf{x}_i\}_{i=1}^N$ as random.

Then the problem of minimizing (2) reduces to the minimization of the *expected risk*

$$R_{\text{exp}} = \frac{1}{2} \int |I - \mathcal{I}| dP(\mathbf{x}) = \frac{1}{2} \int |I - \mathcal{I}| p(\mathbf{x}) d\mathbf{x}. \tag{4}$$

Rather than attempting to estimate probability distributions, such as $p(\mathbf{x})$, from spatially distributed data, statistical learning replaces the expected risk R_{exp} with the *empirical risk*

$$R_{\text{emp}} = \frac{1}{2N} \sum_{i=1}^{N} |I(\mathbf{x}_i) - \mathcal{I}(\mathbf{x}_i)|. \tag{5}$$

These two quantities are related by a probabilistic bound, $R_{\text{exp}} \leq R_{\text{emp}} + C$, where the function C depends on the *Vapnik - Chervonenkis (VC) dimension* and the number of data points N [10][11, Ch. 4]. The VC dimension represents a measure of the complexity of the family of functions $\mathcal{I}$. Analysis of the tightness of this bound, which while providing a useful theoretical motivation for the SVM described below, is often too loose to be of much practical significance, is an active area of research in the field of statistical learning.

III. Support Vector Machines

The SVM is a relatively recent technique that has attracted a great deal of interest due to its excellent performance on a wide range of classification problems, e.g., [11], [10], [12]. The theoretical foundation of this technique is grounded in the fact that the maximal margin SVM, which we describe below, provides a bound of the expected risk R_{exp} [11, Ch. 6, Remark 6.7], [13, Chapter 7].

The simplest maximal margin SVM deals with linearly separable data collected from perfectly stratified geologic media, where different geologic facies are separated by planes (in three dimensions) or straight lines (in two dimensions). It is generalized to accommodate arbitrary data sets by means of Kernel SVMs. Both SVMs and Kernel SVMs are described below.

A. Linearly separable data

Consider a boundary given by the straight line, with equation

$$\mathbf{a} \cdot \mathbf{x} + b = 0. \tag{6}$$

Our goal is to determine the unknown coefficients $\mathbf{a} = (a_1, a_2)^T$ and b from the data set $\{\mathbf{x}_i, I(\mathbf{x}_i)\}_{i=1}^N$. In machine learning, an algorithm for constructing such a boundary between samples from two classes is known as a linear classifier.

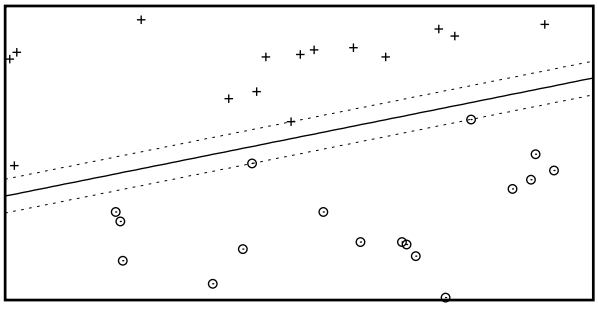

Fig. 1. A schematic representation of the boundary between two heterogeneous geologic facies M_1 and M_2 (located above and below the boundary, respectively) in a perfectly stratified geologic formation. The $+$ and $\odot$ signs indicate the locations where a parameter K is sampled and $I = 1$ or $= 0$, respectively. A maximum margin linear classifier for displayed samples consists of the decision boundary or the boundary estimate (solid line), and the margin (dotted lines).

A maximum margin linear classifier is illustrated in Fig. 1 — the boundary estimate is indicated with the solid line, and

the dotted lines indicate the extent of the *margin*, i.e., the region within which the boundary could be shifted orthogonally without misclassifying any of the data samples. If d_1 and d_2 designate the perpendicular distances from the estimated boundary (solid line) to the nearest data point(s) in materials M_1 and M_2, respectively, then the size of the margin (dotted lines) is $d = d_1 + d_2$, and the sample points determining the position of the margin are called the *support vectors*. Since the lines bounding the margin are parallel to the boundary (6), their normal is also $\mathbf{a}$. The maximal margin SVM determines the coefficients $\mathbf{a}$ and b in (6) by maximizing the size of this margin. While any choice of straight line that lies within the margin provides the same empirical risk R_{emp}, the maximum margin straight line is a principled choice for minimizing the expected risk R_{exp} [13, Section 7.2].

The maximal margin SVM is constructed as follows. Let $\mathbf{a}\cdot\mathbf{x}+b = \pm 1$ be two equations for the dashed lines bounding the margin in Fig. 1. Since the margin separates the two materials, all data points satisfy either

$$\mathbf{a}\cdot\mathbf{x}_i + b \geq +1 \tag{7a}$$

or

$$\mathbf{a}\cdot\mathbf{x}_i + b \leq -1. \tag{7b}$$

Defining the indicator function $J(\mathbf{x}) = 2I(\mathbf{x}) - 1$, so that $J(\mathbf{x}) = -1$ whenever $I(\mathbf{x}) = 0$ and $J(\mathbf{x}) = 1$ whenever $I(\mathbf{x}) = 1$, and denoting $J_i = J(\mathbf{x}_i)$ allows one to combine the two inequalities (7) into one,

$$(\mathbf{a}\cdot\mathbf{x}_i + b)J_i \geq 1 \qquad \text{for} \qquad i \in \{1,\ldots,N\}. \tag{8}$$

The inequalities (8) become equalities for the $\mathbf{x}_i$ that are support vectors. Let $||\mathbf{a}|| \equiv \sqrt{a_1^2 + a_2^2}$ denote the Euclidean length of $\mathbf{a}$. Since the distances ρ_1 and ρ_2 from the coordinate origin to the lines $\mathbf{a}\cdot\mathbf{x}_i + b = 1$ and $\mathbf{a}\cdot\mathbf{x}_i + b = -1$ are, respectively,

$$\rho_1 = -\frac{b+1}{||\mathbf{a}||} \qquad \text{and} \qquad \rho_2 = -\frac{b-1}{||\mathbf{a}||}, \tag{9}$$

the distance between these two lines $\rho_2 - \rho_1$, i.e., the margin d, is given by

$$d = \frac{2}{\|\mathbf{a}\|}. \tag{10}$$

Thus the SVM can be formulated as a problem of maximizing d (or, equivalently, minimizing $||\mathbf{a}||$) subject to the linear constraints (8). Introducing Lagrange multipliers $\gamma_i \geq 0$ for $i \in \{1,\ldots,N\}$ leads to the Lagrangian

$$L(\mathbf{a}, b, \boldsymbol{\gamma}) = \frac{1}{2}||\mathbf{a}||^2 - \sum_{i=1}^{N}\gamma_i\big[(\mathbf{a}\cdot\mathbf{x}_i + b)J_i - 1\big]. \tag{11}$$

A solution of this optimization problem defines $\mathbf{a}$ and b and thus, in accordance with (6), the boundary between the two layers, located at the center of the margin.

Often, the data are not perfectly linearly separable. A more general SVM formulation introduces slack variables $\xi_i \geq 0$ into the optimization, allowing for misclassification but penalizing the sum of the classification errors, so that the problem becomes the minimization of

$$\frac{1}{2}||\mathbf{a}||^2 + C\sum_{i=1}^{N}\xi_i \tag{12}$$

subject to the constraints [12, Ch. 2] [14, Section 12.2.1]

$$(\mathbf{a}\cdot\mathbf{x}_i + b)J_i \geq 1 - \xi_i \qquad \text{for} \qquad i \in \{1,\ldots,N\}. \tag{13}$$

As before, introducing Lagrange multipliers $\gamma_i, \delta_i \geq 0$ for $i \in \{1,\ldots,N\}$ gives the Lagrangian

$$L(\mathbf{a}, b, \boldsymbol{\xi}, \boldsymbol{\gamma}, \boldsymbol{\delta}) = \frac{1}{2}||\mathbf{a}||^2 - \sum_{i=1}^{N}\gamma_i\big[(\mathbf{a}\cdot\mathbf{x}_i + b)J_i - 1 + \xi_i\big] + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\delta_i\xi_i. \tag{14}$$

Denoting the optimal values of $\mathbf{a}$ and b by $\mathbf{a}^\star$ and $b^\star$ respectively, the indicator function $J(\mathbf{x})$ is given by the *decision function*

$$f(\mathbf{x}) = \text{sign}(\mathbf{a}^\star\cdot\mathbf{x} + b^\star). \tag{15}$$

To obtain $\mathbf{a}^\star$ and $b^\star$, it is often convenient to use the dual optimization problem [14, Section 12.2.1]

$$\max_{\boldsymbol{\gamma}}\left\{\sum_{i=1}^{N}\gamma_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\gamma_i\gamma_j J_i J_j[\mathbf{x}_i\cdot\mathbf{x}_j]\right\} \tag{16}$$

subject to the constraints

$$0 \leq \gamma_i \leq C \quad \text{and} \quad \sum_{i=1}^{N}\gamma_i J_i = 0. \tag{17}$$

Let $\gamma_i^\star$ ($i \in \{1,\ldots,N\}$) be solutions of (16) – (17). Then finding the solutions of $\partial L/\partial a_k = 0$ for (14) gives

$$\mathbf{a}^\star = \sum_{i=1}^{N}\gamma_i^\star J_i\mathbf{x}_i. \tag{18}$$

Let $\mathbf{x}_+$ and $\mathbf{x}_-$ denote arbitrary support vectors for which $J = 1$ and $J = -1$, respectively. Then the constraints (13), which at these points become equalities, give

$$b^\star = -\frac{1}{2}\mathbf{a}^\star\cdot(\mathbf{x}_+ + \mathbf{x}_-). \tag{19}$$

Thus a solution for the decision function (15), and hence the indicator function, is

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{N}\gamma_i^\star J_i\mathbf{x}_i\cdot\mathbf{x} + b^\star\right). \tag{20}$$

B. Kernel SVMs

In most practical problems, boundaries between geologic facies are significantly more complex than a straight line or a plane. To account for this geometric complexity, one can generalize the linear maximum margin SVM by noting that data which cannot be separated by a straight line or plane in the two- or three-dimensional space of observation often

become linearly separable (by a hyperplane) when projected onto another, usually higher-dimensional space.

Let $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}^m$ be a mapping of the n-dimensional physical space onto an m-dimensional space (known as a feature space) in which the linear SVM can be applied. An equation for a hyperplane separating the two materials in the m-dimensional space is

$$\mathbf{a} \cdot \mathcal{F}(\mathbf{x}) + b = 0, \tag{21}$$

where parameters $\mathbf{a} \in \mathbb{R}^m$ and b are determined from the transformed data set $\{\mathcal{F}(\mathbf{x}_i), J_i\}_{i=1}^N$ by solving the quadratic optimization of the linear SVM (14),

$$L(\mathbf{a}, b, \boldsymbol{\xi}, \boldsymbol{\gamma}, \boldsymbol{\delta}) = \frac{1}{2}||\mathbf{a}||^2 - \sum_{i=1}^{N} \gamma_i \big[(\mathbf{a} \cdot \mathcal{F}(\mathbf{x}_i) + b) J_i - 1 + \xi_i)\big] + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \delta_i \xi_i. \tag{22}$$

In analogy to (15), the decision function is given by

$$f(\mathbf{x}) = \text{sign}(\mathbf{a}^\star \cdot \mathcal{F}(\mathbf{x}) + b^\star). \tag{23}$$

It is linear in the feature space, is nonlinear in the physical (or input) space.

The dual optimization problem (16) is now recast as

$$\max_{\boldsymbol{\gamma}} \left\{ \sum_{i=1}^{N} \gamma_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_i \gamma_j J_i J_j [\mathcal{F}(\mathbf{x}_i) \cdot \mathcal{F}(\mathbf{x}_j)] \right\} \tag{24}$$

subject to the constraints

$$0 \le \gamma_i \le C \quad \text{and} \quad \sum_{i=1}^{N} \gamma_i J_i = 0. \tag{25}$$

The key observation here is that the feature space vectors enter into the optimization only within an inner product. If a kernel function

$$\mathcal{K}(\mathbf{x}, \mathbf{x}') = \mathcal{F}(\mathbf{x}) \cdot \mathcal{F}(\mathbf{x}') \tag{26}$$

is available for a specific mapping $\mathcal{K}$, the required inner products may be computed directly from the physical space, without explicitly performing the potentially computationally expensive mapping into the feature space. Hence the dual optimization (24) may be expressed as

$$\max_{\boldsymbol{\gamma}} \left\{ \sum_{i=1}^{N} \gamma_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \gamma_i \gamma_j J_i J_j \mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) \right\}, \tag{27}$$

avoiding explicit computation of the mapping $\mathcal{F}$.

In analogy to (18) – (20), the decision function f is now given by

$$f(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{N} \gamma_i^\star J_i \mathcal{K}(\mathbf{x}, \mathbf{x}_i) + b^\star \right), \tag{28}$$

with $\gamma_i^\star$ ($i \in \{1, \ldots, N\}$) defined as a solution of the dual optimization problem (27),

$$\mathbf{a}^\star = \sum_{i=1}^{N} \gamma_i^\star J_i \mathcal{F}(\mathbf{x}_i) \tag{29}$$

and $b^\star$ given by (19). Note that the decision function (28) is expressed in terms of the kernel $\mathcal{K}$, without the need for explicit mapping onto the feature space.

Among a wide variety of Mercer kernels, we will consider the performance of the polynomial kernel of order p

$$\mathcal{K}_{\text{PLM}}(\mathbf{x}, \mathbf{x}') = (\mathbf{x} \cdot \mathbf{x}' + 1)^p, \tag{30a}$$

and the sigmoid kernel

$$\mathcal{K}_{\text{SIG}}(\mathbf{x}, \mathbf{x}') = \tanh(\rho \mathbf{x} \cdot \mathbf{x}' + \varrho), \tag{30b}$$

the exponential radial basis function kernel

$$\mathcal{K}_{\text{ERB}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{||\mathbf{x} - \mathbf{x}'||}{2\sigma^2} \right), \tag{31}$$

and the Gaussian radial basis function kernel

$$\mathcal{K}_{\text{GRB}}(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{||\mathbf{x} - \mathbf{x}'||^2}{2\sigma^2} \right). \tag{32}$$

IV. Synthetic Example

To demonstrate the applicability of SVMs to subsurface imaging, and to elucidate its relative advantages with respect to a geostatistical approach, we reconstruct, from a few data points selected at random from a uniform distribution, the boundaries between two heterogeneous geologic facies in a synthetic porous medium shown in Fig. 2. This synthetic example was generated as follows.

We start by generating two autocorrelated, weakly stationary, and normally distributed processes, representing two distinct spatial distributions of log hydraulic conductivity $Y = \ln K$ with the ensemble means of -0.1 and 7.0. When hydraulic conductivities are expressed in $[cm/day]$, this corresponds to clayey and sandy materials, respectively. Both log-conductivity distributions have unit variance and Gaussian autocorrelation with unit correlation scale. We take these two fields to be mutually uncorrelated, although the methodology is capable of taking this feature into account. The fields are generated by the SGSIM code [15] on a 60×60 grid, using a grid spacing of $1/5$ of the log-conductivity correlation length.

Next, the composite porous medium in Fig. 2 is constructed by setting an arbitrary shape of the internal boundary between the two materials and by assigning values of log-conductivity to cells in the domain. Assigning the indicator function (1) with a threshold value of 4.0 to each element on the grid results in Fig. 3.

We use an SVM [12] to reconstruct the boundary between the two geologic facies in Fig. 3 from a few (randomly) selected data points. Sampling densities ranging from 0.25% (9 data points) to 20% (720 data points) were considered. Fig. 4 compares the performance of the SVMs whose kernels are given by the polynomial (PLM), exponential radial basis (ERB), Gaussian radial basis (GRB), and sigmoid (SIG) functions in (30). For each sampling density, we randomly generated 20 realizations of the locations of data points and counted the number of elements on the grid that were misclassified by the SVMs. The error in Fig. 4 represents the average (over 20 realizations) number of misclassified elements. One can see that the kernels given by the exponential radial basis (ERB)

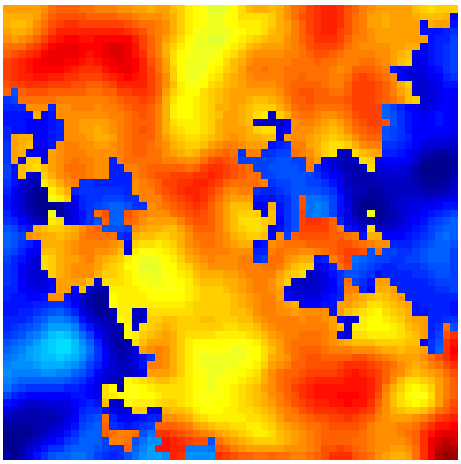

Fig. 2. Synthetic data on a 60×60 grid. Values range between -2.04 and 9.89.

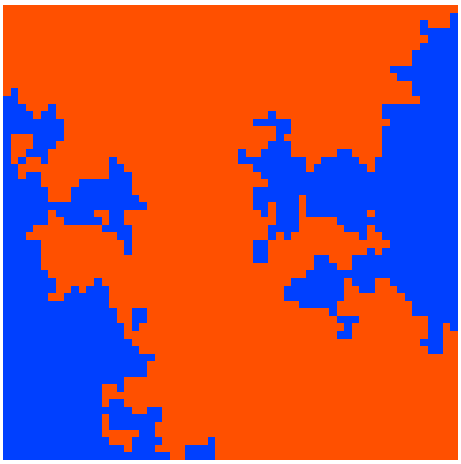

Fig. 3. Classification of data in Figure 2, obtained by setting a threshold value of 4.0.

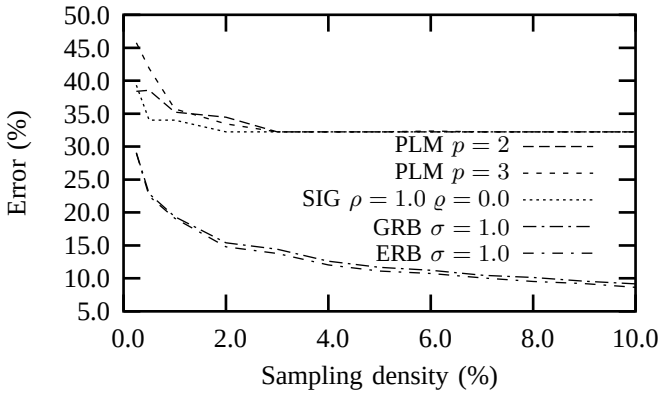

Fig. 4. Error rates corresponding to the SVMs with polynomial (PLM), exponential radial basis (ERB), and Gaussian radial basis (GRB) kernels.

functions perform best, which is compatible with findings in a wide variety of other SVM applications.

Figs. 6 and 5 demonstrate the sensitivity to a fitting parameter σ of the SVMs with the exponential radial basis (ERB) and Gaussian radial basis (GRB) kernels given by (31) and (32), respectively. One can see that the performance of both SVMs is relatively insensitive to the choice of σ (with its values varying over about two orders of magnitude), when the sampling density exceeds 2%. This finding is encouraging, since the optimal choice of σ is nontrivial.

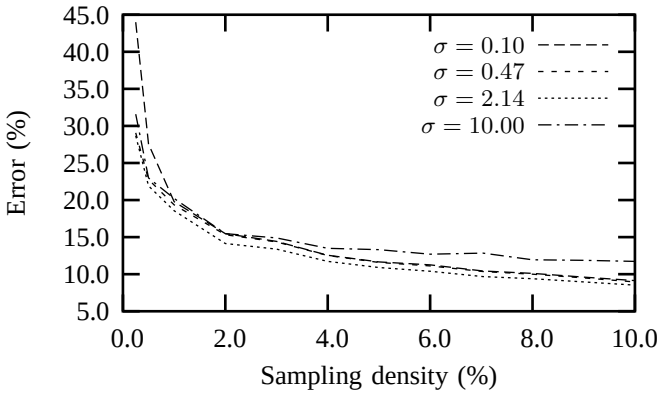

Fig. 5. Error rates corresponding to SVMs with an ERB kernel (31) and several values of σ.

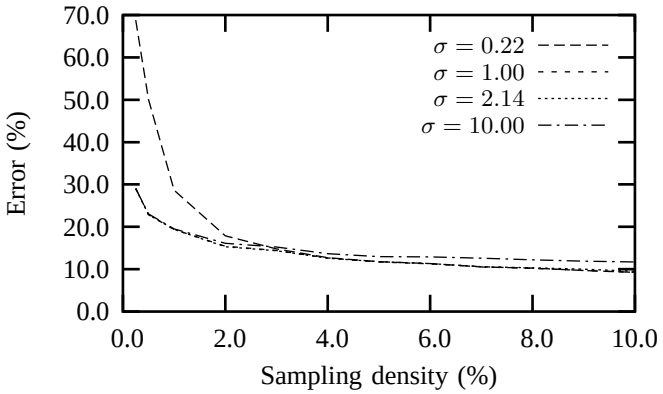

Fig. 6. Error rates corresponding to SVMs with an GRB kernel (32) and several values of σ.

Figs. 7 and 8 show the geologic facies reconstructed by an ERB SVM with $\sigma = 1.0$ from 9 and 180 sample points, respectively. The locations of sample points are indicated by the lighter shades. The comparison of these reconstructions with the true field in Fig. 3 shows that even very sparse sampling might be sufficient for the SVMs to capture general trends in the spatial arrangement of geologic facies. However, the performance of the SVMs on such sparse data sets is highly dependent (i.e., highly variable from one realization to another) on the actual locations of data points. As the sampling density increases, the SVMs capture finer features of the spatial arrangement of geologic facies, and their performance is less dependent on a sampling realization.

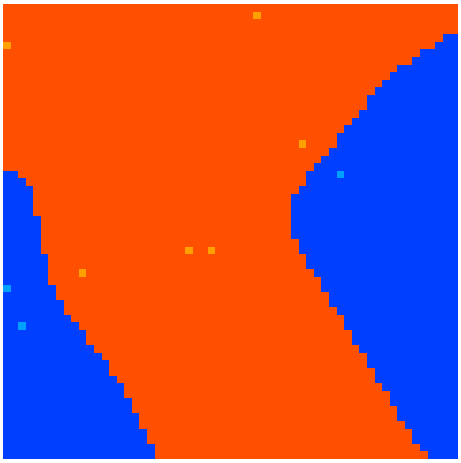

Fig. 7. Classification of data in Fig. 2, obtained by an ERB SVM using 9 sample points (0.25% sampling density).

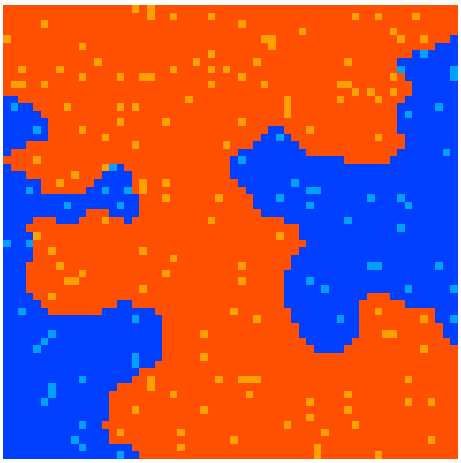

Fig. 8. Classification of data in Fig. 2, obtained by an ERB SVM using 180 sample points (5% sampling density).

A. Comparison with a geostatistical approach

We compare the accuracy of the facies reconstruction by means of the SVM with that obtained by a geostatistical approach (GSA) described in the Appendix. It is important to note that this and other geostatistical approaches to facies delineation assume that the relative volumes occupied by the two materials obtained from a sample are representative of the whole field. This assumption is usually difficult to validate a priori.

Fig. 9 shows the comparison of the performance of the GSA and the SVM with the ERB kernel and $\sigma = 1.0$ consisted of 20 trials for each of sampling densities. When enough measurements are available (i.e., when the sampling density is large enough), both methods perform equally well, with the SVM being slightly more accurate than the GSA. Two factors, however, argue strongly in favor of SVMs. First, they perform relatively well even on highly sparse data sets (see the boundary reconstruction from 9 sampling points in Fig. 7), on which GSA fails. Second, SVMs are highly automated, while GSAs require manual data analysis to construct spatial variograms. As a result, the GSAs are highly time consuming and depend on the subjective judgement of the practitioner.

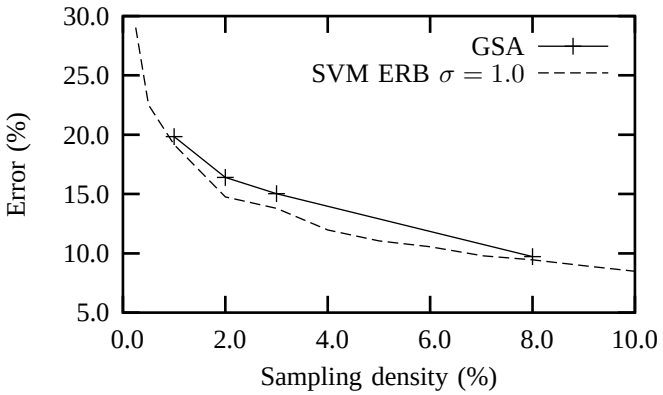

Fig. 9. Error rates corresponding to the GSA and SVM approaches.

V. Support Vector Regression

In hydrologic applications, the delineation of geologic facies from parameter data is often not sufficient. Since many geologic facies are heterogeneous, it is also necessary to assign parameter values to locations (e.g., elements of a numerical grid) where data are not available. Geostatistical approaches achieve this goal through *data interpolation* algorithms, such as Kriging [15]. SVMs take an alternative route by employing *data regression* strategies.

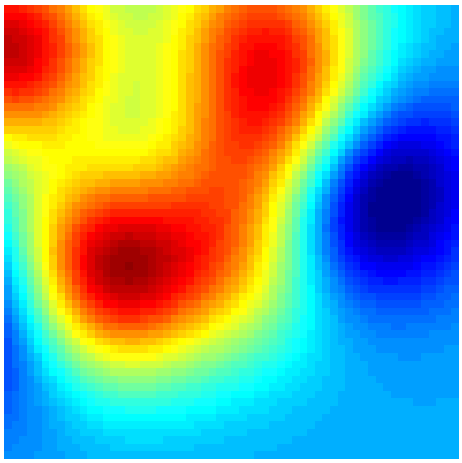

Fig. 10. Regression of 9 sample points (0.25% sampling density) from data in Figure 2, obtained by an GRB ($\sigma = 10.0$) SVM regression.

SVM regression may be viewed as a generalization of the SVM classification introduced in Section II to delineate the boundaries between geologic facies. Let $\hat{K}(\mathbf{x}, \boldsymbol{\alpha})$ be an estimate of a "true" parameter field $K(\mathbf{x})$, whose adjustable parameters $\boldsymbol{\alpha}$ are determined from the available parameter data $\{K_i = K(\mathbf{x}_i)\}_{i=1}^{N}$. SVM regression aims to minimize the difference between the two, while providing a probabilistic bound on the accuracy of the estimator $\hat{K}$ at a randomly drawn point $\mathbf{x}$ [11, Sec. 4.5].

Similar to SVM classification in Section III, SVM regression is first introduced for a linear regression and is then generalized for a nonlinear regression through the use of

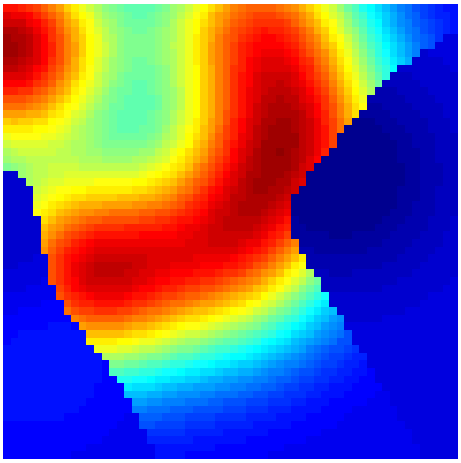

Fig. 11. Regression of 9 sample points (0.25% sampling density) from data in Figure 2, obtained by an ERB ($\sigma = 1.0$) SVM classification followed by an GRB ($\sigma = 10.0$) SVM regression.

kernels. For linear regression, we seek to approximate a data set $\{K_i = K(\mathbf{x}_i)\}_{i=1}^N$ with a linear function

$$\hat{K}(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x} + b. \tag{33}$$

The regression equivalent of the classification functional (12) is [11, Sec. 6.2]

$$\frac{1}{2}||\mathbf{a}||^2 + C\sum_{i=1}^{N} L_\epsilon(\mathbf{x}_i, K_i, \hat{K}), \tag{34}$$

where the pointwise sum of classification errors $\sum_{i=1}^N \xi_i$ in (12) is replaced by pointwise sum $\sum_{i=1}^N L_\epsilon(\mathbf{x}_i, K_i, \hat{K})$ of a *loss function* $L(\mathbf{x}, K, \hat{K})$ which measures the error in approximating K_i at $\mathbf{x}_i$ by $\hat{K}(\mathbf{x}_i)$. While it is not the only choice, we utilize the ϵ-insensitive loss function

$$L_\epsilon(\mathbf{x}, K, \hat{K}) = \begin{cases} 0 & \text{for } |\hat{K}(\mathbf{x}) - K| < \epsilon \\ |\hat{K}(\mathbf{x}) - K| - \epsilon & \text{otherwise.} \end{cases} \tag{35}$$

first used in SVM regression. With this loss function, the primal optimization problem for SVM regression is the minimization of

$$\frac{1}{2}||\mathbf{a}||^2 + C\sum_{i=1}^{N}(\xi_i + \hat{\xi}_i) \tag{36}$$

subject to the constraints

$$(\mathbf{a} \cdot \mathbf{x}_i + b) - K(\mathbf{x}_i) \le \epsilon + \xi_i \tag{37a}$$

$$K(\mathbf{x}_i) - (\mathbf{a} \cdot \mathbf{x}_i + b) \le \epsilon + \hat{\xi}_i, \tag{37b}$$

where $i \in \{1, \ldots, N\}$ and $\xi_i \ge 0,\ \hat{\xi}_i \ge 0$.

Introducing Lagrange multipliers $\gamma_i, \hat{\gamma}_i, \delta_i, \hat{\delta}_i \ge 0$ ($i \in \{1, \ldots, N\}$) gives the Lagrangian

$$\begin{aligned} L(\mathbf{a}, b, \boldsymbol{\xi}, \hat{\boldsymbol{\xi}}, \boldsymbol{\gamma}, \hat{\boldsymbol{\gamma}}, \boldsymbol{\delta}, \hat{\boldsymbol{\delta}}) = \frac{1}{2}||\mathbf{a}||^2 + C\sum_{i=1}^{N}(\xi_i + \hat{\xi}_i) \\ - \sum_{i=1}^{N}\gamma_i(\epsilon + \xi_i + K_i - \mathbf{a}\cdot\mathbf{x}_i - b) \\ - \sum_{i=1}^{N}\hat{\gamma}_i(\epsilon + \hat{\xi}_i - K_i + \mathbf{a}\cdot\mathbf{x}_i + b) \\ - \sum_{i=1}^{N}\delta_i\xi_i - \sum_{i=1}^{N}\hat{\delta}_i\hat{\xi}_i. \end{aligned} \tag{38}$$

In analogy with (16), the dual optimization problem is

$$\begin{aligned} \max_{\boldsymbol{\gamma}, \hat{\boldsymbol{\gamma}}} \Bigg\{ \sum_{i=1}^{N}(\gamma_i - \hat{\gamma}_i)K_i - \epsilon\sum_{i=1}^{N}(\gamma_i + \hat{\gamma}_i) \\ -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\gamma_i - \hat{\gamma}_i)(\gamma_j - \hat{\gamma}_j)[\mathbf{x}_i \cdot \mathbf{x}_j] \Bigg\} \end{aligned} \tag{39}$$

subject to the constraints

$$0 \le \gamma_i \le C, \qquad 0 \le \hat{\gamma}_i \le C \qquad \sum_{i=1}^{N}(\gamma_i - \hat{\gamma}_i) = 0. \tag{40}$$

Let $\gamma_i^\star$ and $\hat{\gamma}_i^\star$ ($i \in \{1, \ldots, N\}$) denote a solution of (39) – (40). Then the optimal parameters $\mathbf{a}$ and b are given by [13, Section 9.2]

$$\mathbf{a}^\star = \sum_{i=1}^{N}(\gamma_i^\star - \hat{\gamma}_i^\star)\mathbf{x}_i. \tag{41}$$

and

$$b^\star = K_j - \mathbf{a}^\star \cdot \mathbf{x}_j - \epsilon, \tag{42}$$

respectively, where K_j and $\mathbf{x}_j$ in (42) are chosen such that $0 < \gamma_j^\star < C$.

For a non-linear regression, we once again replace $\mathbf{x}$ by $\mathcal{F}(\mathbf{x})$ to give the regression

$$\hat{K}(\mathbf{x}) = \mathbf{a} \cdot \mathcal{F}(\mathbf{x}) + b,$$

and utilize the dual form of the optimization to express the problem in terms of the kernel $\mathcal{K}$ associated with mapping $\mathcal{F}$, giving the regression

$$\hat{K}(\mathbf{x}) = \sum_{i=1}^{N}(\gamma_i - \hat{\gamma}_i)\mathcal{K}(\mathbf{x}_i, \mathbf{x}) + b. \tag{43}$$

In addition to the direct use of the nonlinear regression (43), we explore a two-step procedure. First, we use the SVM to delineate the geologic facies from a data set $\{K(\mathbf{x}_i)\}_{i=1}^N$. Then we perform a nonlinear regression (43) on the data subsets within each facies separately. In the simulations presented here, we used the regression SVM [12] with the GBF kernel (32) and $\sigma = 1$ for the direct data regression; and the classification SVM [12] with the GRB kernel (31) and $\sigma = 1$ and the regression SVM [12] with the GBF kernel (32) and $\sigma = 1$ for the sequential data regression.

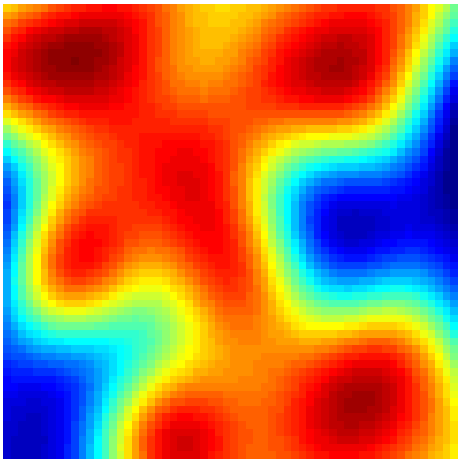

Fig. 12. Regression of 180 sample points (5% sampling density) from data in Figure 2, obtained by an GRB ($\sigma = 10.0$) SVM regression.

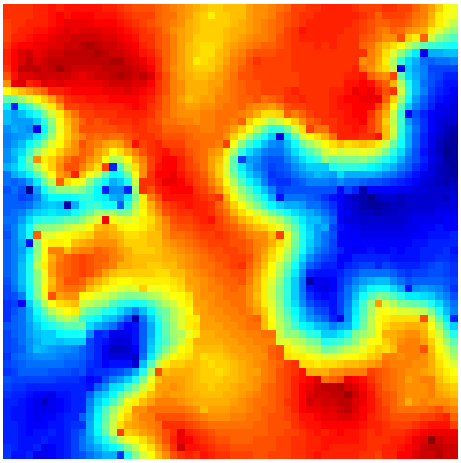

Fig. 14. Interpolation of 180 sample points (5% sampling density) from data in Figure 2, obtained by direct Kriging.

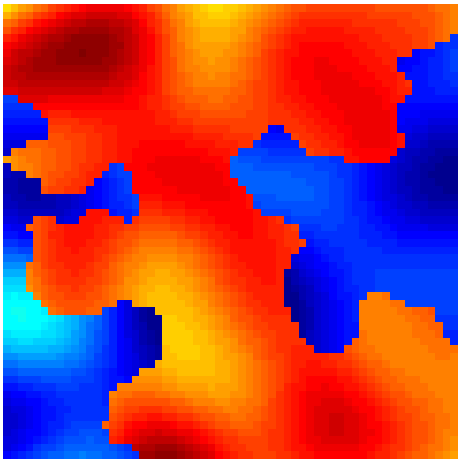

Fig. 13. Regression of 180 sample points (5% sampling density) from data in Figure 2, obtained by an ERB ($\sigma = 1.0$) SVM classification followed by an GRB ($\sigma = 10.0$) SVM regression.

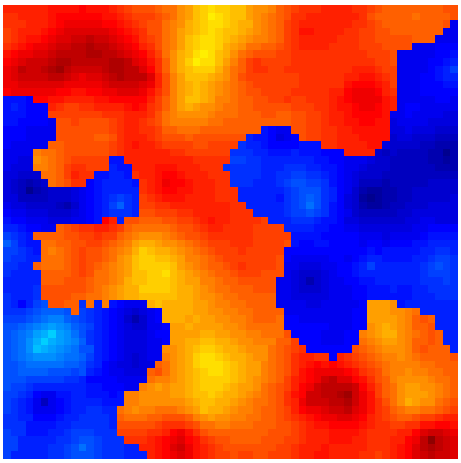

Fig. 15. Interpolation of 180 sample points (5% sampling density) from data in Figure 2, obtained by Kriging with classification.

Figs. 10 and 11 compare the performance of the two regression strategies for a sparse data set (9 data points resulting in the sampling density of 0.25%) shown in Fig. 7 by the pale squares. Figs. 12 and 13 provide a similar comparison for a denser data set (180 data points, corresponding to a sampling density of 5%) shown in Fig. 8. As one would expect, this comparison shows that the quality of the reconstruction of the K field increases with the sampling density. The proposed two-step SVM regression (Figs. 11 and 13) outperforms the direct SVM regression (Figs. 10 and 12), capturing some of the main features of the K distribution even from an extremely sparse (the sampling density of 0.25%) data set.

Finally, we compare the SVM regressions with two geostatistical approaches, which employ alternative interpolation strategies based on Kriging. The first approach uses simple Kriging to interpolate between the 180 data points shown in Fig. 8. An isotropic spherical variogram provides the best fit of the data with the following parameters: nugget = 1.08, sill = 12.01, and range = 5.261. Note that the synthetic data set in Figs. 2 and 3 is a realization of the random field with a Gaussian variogram. The discrepancy between the variogram fitted to the 180 data points and the variogram used to construct an underlying reference field is due to the finite number of samples and their spatial location.

The second approach, which we call Kriging with classification, consists of two steps. First, we use the geostatistical facies delineation procedure described in the Appendix. Second, we use simple Kriging on the two subsets of the 180 data set, each of which belongs to one of the two facies. This procedure results in an isotropic exponential variogram with nugget 0.0, sill 0.7 and range 3.357 for the high conductivity facies (the red portion of the domain in Fig. 8), and in an isotropic exponential variogram with nugget 0.035, sill 0.7 and range 1.768 for the low conductivity facies (the blue portion of the domain in Fig. 8).

Figs. 14 and 15 provide the K fields reconstructed with the simple Kriging and the Kriging with classification, respectively. The Kriging with classification outperforms the simple Kriging, as is the case with their SVM counterparts. The comparison of the two SVM approaches (Figs. 12 and 13) with the two Kriging approaches (Figs. 14 and 15) gives a slight edge to the latter, at least when an eyeball measure is used. At the same time, it is worthwhile to point out that the

Kriging approaches (especially the Kriging with classification) are much more labor intensive than their SVM counterparts.

VI. Conclusions

We explored the potential of the statistical learning theory in general, and support vector machines (SVMs) in particular, to delineate geologic facies from limited data. This was accomplished (i) by reconstructing, from a few data points, a synthetic randomly generated porous medium consisting of two heterogeneous materials; and (ii) by comparing the performance of the SVMs with that of the geostatistical approach [16].

Key differences between the SVMs and geostatistics, first pointed out in [8], are

- Since the SVMs do not treat the subsurface environment as random, they do not require ergodicity and other statistical assumptions that lie at the heart of geostatistics;
- While geostatistics provides a set of interpolation tools, the SVMs use regression.

Our analysis leads to the following major conclusions:

- For any sampling densities the SVMs slightly outperforms the geostatistical approach in reconstructing the boundaries between two geologic facies, while significantly reducing the computational time;
- For very low sampling densities (e.g., 0.25%), which make the inference of statistical parameters meaningless, the geostatistical approach fails, while the SVMs still do a reasonably good job in reconstructing the boundaries.

We also employed the SVMs and geostatistics to infer parameter values at spatial locations where parameter data are not not available. We found that

- sequential, two-step regression SVMs (geostatistical approach), which first use SVMs (Indicator Kriging) to delineate geologic facies and then use SVM regression (Kriging interpolation), outperform the direct application of the regression SVMs (Kriging);
- the reconstructions of parameter fields obtained with the two approaches are comparable at medium to high sampling density;
- regression SVMs are significantly less expensive computationally and require less data.

Acknowledgment

This research was performed under the auspices of the U.S. Department of Energy, under contract W-7405-ENG-36. This work was supported in part by the U.S. Department of Energy under the DOE/BES Program in the Applied Mathematical Sciences, Contract KC-07-01-01, and in part by the LDRD Program at Los Alamos National Laboratory. This work made use of shared facilities supported by SAHRA (Sustainability of semi-Arid Hydrology and Riparian Areas) under the STC Program of the National Science Foundation under agreement EAR-9876800.

Appendix

To evaluate the relative strengths of Kernel SVMs, we compare their performance with that the geostatistical approach due to [16]. This approach consists of the following steps: First, we use Kriging [15] to construct a map of the ensemble average of the indicator function $\langle I(\mathbf{x})\rangle$ from the data $\{I(\mathbf{x}_i)\}_{i=1}^N$. The ensemble mean $I(\mathbf{x})$ is the probability that a point $\mathbf{x}$ lies in Material 1, $\langle I(\mathbf{x})\rangle = P[\mathbf{x} \in M_1]$. Then we define a boundary between the two materials as an isoline $P[\mathbf{x} \in M_1] = c$, where c is a number of data points in Material 1 (or 2) relative to the total number of data points, after accounting for data clustering.

In some cases, this value does not guarantee that the Kriging estimate of the fraction of the total area covered by the low-conductivity material equals the declustered global mean of the original indicator data, resulting from the raw data. In such cases, c is set to a value of the Kriged indicator field which allows one to recover a reconstruction that honors the empirical relative volumetric fractions of the two materials.

References

[1] Y. Rubin, *Applied stochastic hydrogeology*. New York: Oxford University Press, 2003.

[2] A. G. Journel, "Nonparametric estimation of spatial distribution," *Math. Geol.*, vol. 15, no. 3, pp. 445 – 468, 1983.

[3] A. G. Journel and E. K. Isaaks, "Conditional Indicator simulation: Application to a Saskatcheouan Uranium deposit," *Math. Geol.*, vol. 16, no. 7, pp. 685 – 718, 1984.

[4] A. G. Journel and J. J. Gomez-Hernandez, "Stochastic imaging of the Wilmington clastic sequence," *SPE Form. Eval.*, vol. 8, no. 1, pp. 33 – 40, March 1993.

[5] W. Oh and W. B. Lindquist, "Image thresholding by indicator kriging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 7, pp. 590 – 602, 1999.

[6] C. M. Bishop, *Neural networks for pattern recognition*. Oxford: Oxford University Press, 1995.

[7] S. Moysey, J. Caers, R. Knight, and R. M. Allen-King, "Stochastic estimation of facies using ground penetrating radar data," *Stoch. Environ. Res. Risk Assessm.*, vol. 17, pp. 306 – 318, 2003.

[8] D. M. Tartakovsky and B. E. Wohlberg, "Delineation of geologic facies with statistical learning theory," *Geophys. Res. Lett.*, vol. 31, no. 18, p. L18502, 2004, doi:10.1029/2004GL020864.

[9] V. N. Vapnik, *Statistical Learning Theory*. New York: John Wiley & Sons, Inc., 1998.

[10] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[11] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, UK: Cambridge University Press, 2000.

[12] S. R. Gunn, "Support vector machines for classification and regression," University of Southampton, Southampton, U.K., Technical Report, School of Electronics and Computer Science, 1998. [Online]. Available: http://www.ecs.soton.ac.uk/~srg/publications/pdf/SVM.pdf

[13] B. Schölkopf and A. J. Smola, *Learning with Kernels*. Cambridge, MA, USA: The MIT Press, 2002.

[14] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY, USA: Springer, 2001.

[15] C. V. Deutsch and A. G. Journel, *Geostatistical Software Library and User's Guide*. New York: Oxford University Press, 1992.

[16] R. W. Ritzi, D. F. Jayne, A. J. Z. Jr., A. A. Field, and G. E. Fogg, "Geostatistical modeling of heterogeneity in glaciofluvial, buried-valley aquifer," *Ground Water*, vol. 32, no. 4, pp. 666–674, July – August 1994.

PLACE PHOTO HERE

Brendt Wohlberg received the BSc(Hons) degree in applied mathematics, and the MSc(Applied Science) and PhD degrees in electrical engineering from the University of Cape Town, South Africa, in 1990, 1993 and 1996 respectively. He is currently a technical staff member in the Mathematical Modeling and Analysis Group (T-7) at Los Alamos National Laboratory, Los Alamos, NM. His research interests include image coding, pattern recognition, wavelets and adaptive signal decompositions, and inverse problems in signal and image processing.

PLACE PHOTO HERE

Daniel M. Tartakovsky received the MSc (with honours) degree in applied mathematics and fluid mechanics from Kazan State University, Russia, in 1991, and the PhD in hydrology from the University of Arizona, Tucson, in 1996. He is currently a team leader and a technical staff member in the Mathematical Modeling and Analysis Group (T-7) at Los Alamos National Laboratory, Los Alamos, NM. His research interests include uncertainty quantification, stochastic partial differential equations, and inverse modeling.

PLACE PHOTO HERE

Alberto Guadagnini received his degree in Hydraulic Engineering from the Politecnico di Milano, Milano, Italy, in 1989 and the Doctoral Degree in Hydraulic Engineering from the Politecnico di Milano, Milano, Italy, in 1992. He is currently a full professor of Hydraulics at the Department of Hydraulic, Environmental, Infrastructure and Surveying Engineering of the Politecnico di Milano. His research interests include hydrogeology, fluid mechanics, modeling of contaminant fluxes in porous formations, inverse modeling, and stochastic partial differential equations.